Machine Learning Lab Week 10

Name: Chandan Chatragadda – PES2UG23CS141
Sem 5 – C
Date: 10th Oct 2025

Training SVMs using 3 different kernels: Linear, RBF & Polynomial

Analysis Questions for Moons:

1. Based on the metrics and the visualizations, what inferences about the performance of the Linear Kernel can you draw?
2. Compare the decision boundaries of the RBF and Polynomial kernels. Which one seems to capture the shape of the data more naturally?

ANSWERS:

1. **Performance:** The Linear Kernel on the Moons dataset achieves an accuracy of 0.87. While this is a decent accuracy, it's lower than the RBF and Polynomial kernels.
2. **Decision Boundary:** The visualization shows that the Linear Kernel creates a straight-line decision boundary. This boundary struggles to effectively separate the two interlocking half-moon shapes, which are non-linearly separable. It misclassifies points from both classes that are located in the area where the two moons interlock (The area around the center).

Inference: The Linear Kernel is not well-suited for datasets with non-linear structures like the Moons dataset, as it can only create linear decision boundaries.

Analysis Questions for Banknote:

1. In this case, which kernel appears to be the most effective?
2. The Polynomial kernel shows lower performance here compared to the Moons dataset. What might be the reason for this?

ANSWER:

1. The RBF Kernel appears to be most effective as it has the highest precision, recall and F1-score as compared to Linear and Polynomial kernels. The decision boundary plot also visually suggests a better classification in case of RBF Kernel.
2. The reason for the difference in performance is because of the different way the data is distributed in the datasets. The Moon dataset has a better and a much more simpler non-linear pattern that can be handled by polynomial kernel.While the Banknote dataset has a more complex non-linear structure that the polynomial kernel is not very suited to handle. On the contrary, the RBF Kernel can create a more flexible and localized decision boundaries and hence performs better than the linear and polynomial kernel.

Analysis Questions for Hard and Soft Margins:

1. Compare the two plots. Which model, the "Soft Margin" (C=0.1) or the "Hard Margin" (C=100), produces a wider margin?
2. Look closely at the "Soft Margin" (C=0.1) plot. You'll notice some points are either inside the margin or on the wrong side of the decision boundary. Why does the SVM allow these "mistakes"? What is the primary goal of this model?
3. Which of these two models do you think is more likely to be overfitting to the training data? Explain your reasoning.
4. Imagine you receive a new, unseen data point. Which model do you trust more to classify it correctly? Why? In a real-world scenario where data is often noisy, which value of C (low or high) would you generally prefer to start with?

ANSWER:

1. The Soft Margin produces a wider margin.
2. This is done to avoid overfitting and give us a better generalization on unseen data. Hence, our goal is to find a decision boundary that has the largest possible margin while keeping the number of misclassifications under a tolerable limit.
3. The Hard Margin is more likely to be overfitting. This is because it is trying to perfectly separate all training points (including outliers and noisy data) rather than generalize a decision boundary. This results in a very narrow margin that is highly sensitive and does not generalize well.
4. The Soft Margin is more trustable with real-world data as it is better in generalization and does better with noisy data and outliers (which are present in real-world data). Considering the noisy real world data, it is preferable to start with a low value of C so it allows for a soft margin.

SCREENSHOTS:

MOON DATASET:

Classification Report for SVM with LINEAR Kernel with SRN:

```
SVM with LINEAR Kernel <PES2UG23CS141>
              precision    recall  f1-score   support

           0       0.85      0.89      0.87        75
           1       0.89      0.84      0.86        75

    accuracy                           0.87       150
   macro avg       0.87      0.87      0.87       150
weighted avg       0.87      0.87      0.87       150

----------------------------------------
```

Classification Report for SVM with RBF Kernel with SRN

```
SVM with RBF Kernel <PES2UG23CS141>
              precision    recall  f1-score   support

           0       0.95      1.00      0.97        75
           1       1.00      0.95      0.97        75

    accuracy                           0.97       150
   macro avg       0.97      0.97      0.97       150
weighted avg       0.97      0.97      0.97       150

---------------------------------------
```

Classification Report for SVM with POLY Kernel with SRN

```
SVM with POLY Kernel <PES2UG23CS141>
              precision    recall  f1-score   support

           0       0.85      0.95      0.89        75
           1       0.94      0.83      0.88        75

    accuracy                           0.89       150
   macro avg       0.89      0.89      0.89       150
weighted avg       0.89      0.89      0.89       150

---------------------------------------
```

BANKNOTE DATASET:

Classification Report for SVM with LINEAR Kernel

```
SVM with LINEAR Kernel <PES2UG23CS141>
              precision    recall  f1-score   support

      Forged       0.90      0.88      0.89       229
     Genuine       0.86      0.88      0.87       183

    accuracy                           0.88       412
   macro avg       0.88      0.88      0.88       412
weighted avg       0.88      0.88      0.88       412

---------------------------------------
```

Classification Report for SVM with RBF Kernel

```
SVM with RBF Kernel <PES2UG23CS141>
              precision    recall  f1-score   support

      Forged       0.96      0.91      0.94       229
     Genuine       0.90      0.96      0.93       183

    accuracy                           0.93       412
   macro avg       0.93      0.93      0.93       412
weighted avg       0.93      0.93      0.93       412

----------------------------------------
```

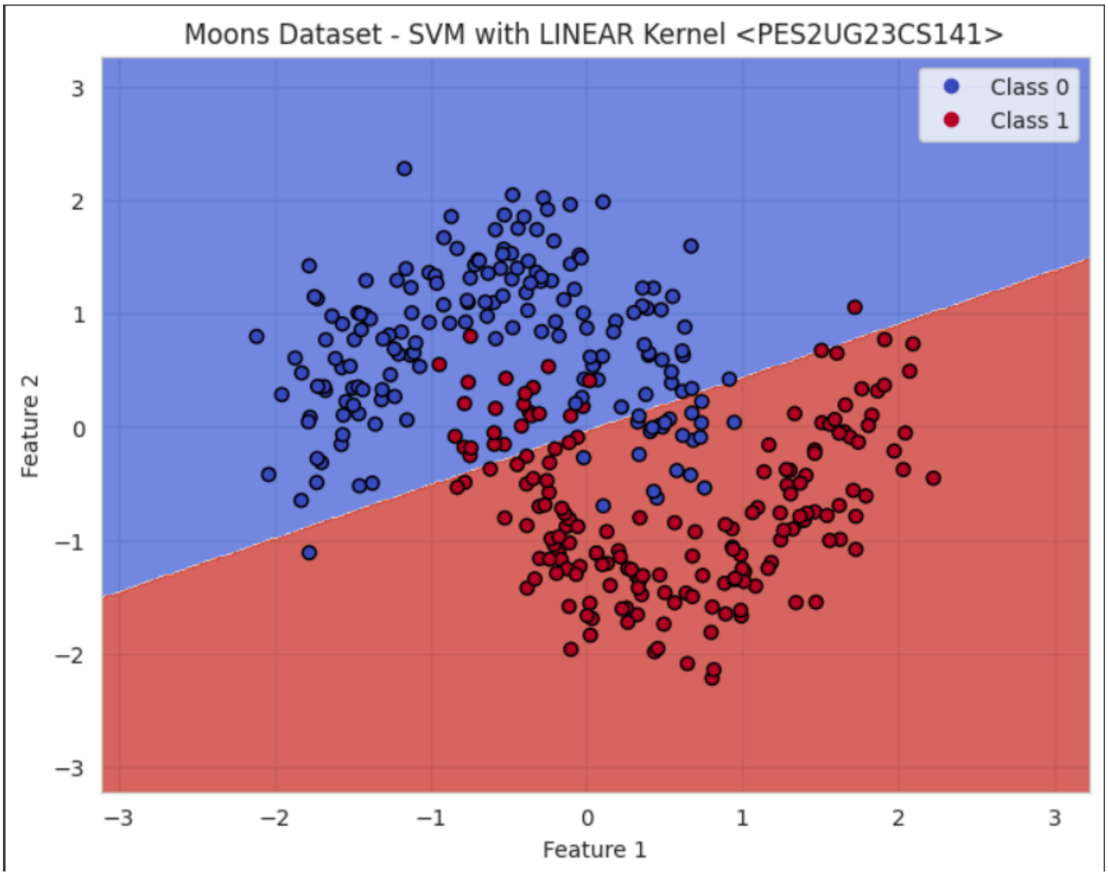
Classification Report for SVM with POLY Kernel

```
SVM with POLY Kernel <PES2UG23CS141>
              precision    recall  f1-score   support

      Forged       0.82      0.91      0.87       229
     Genuine       0.87      0.75      0.81       183

    accuracy                           0.84       412
   macro avg       0.85      0.83      0.84       412
weighted avg       0.85      0.84      0.84       412

----------------------------------------
```

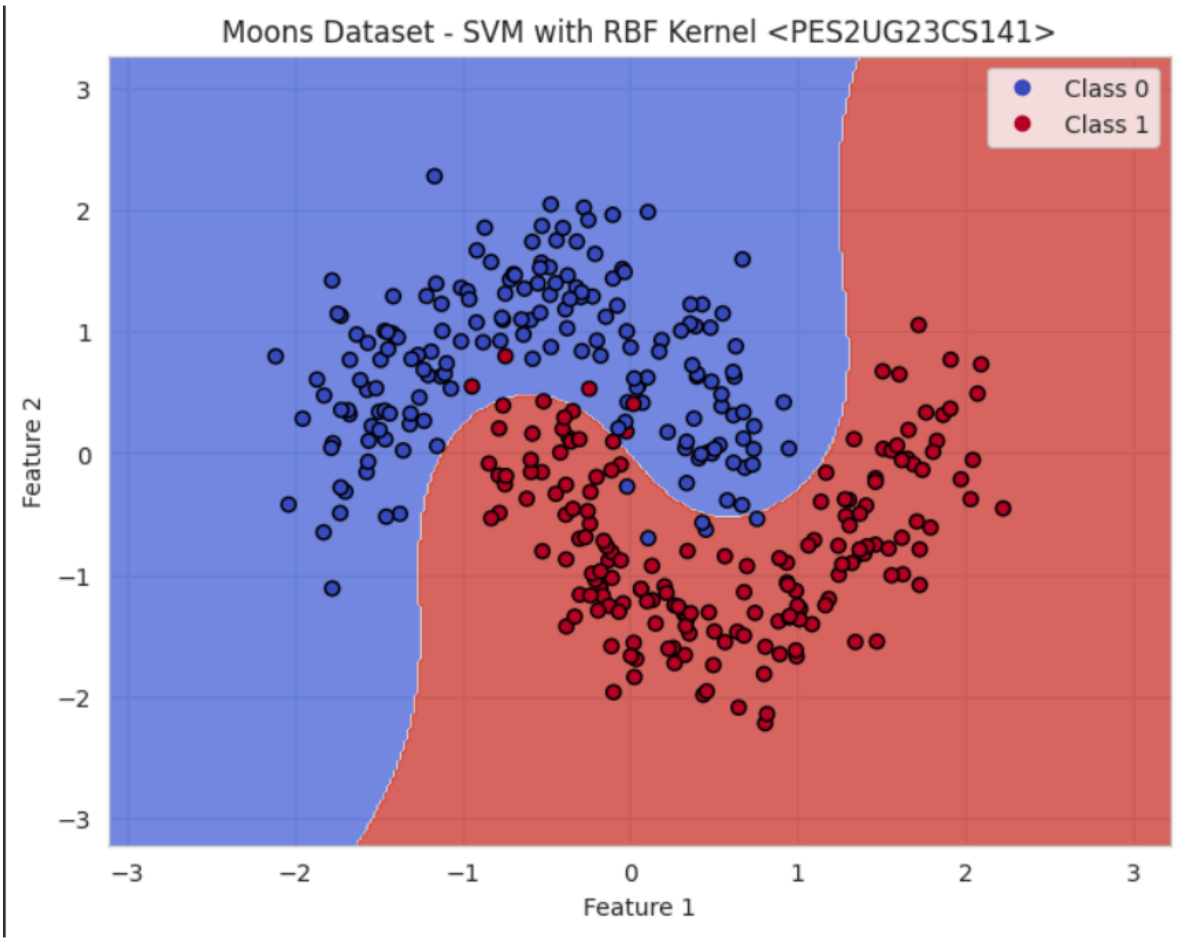
DECISION BOUNDARY VISUALIZATIONS

Moons Dataset:

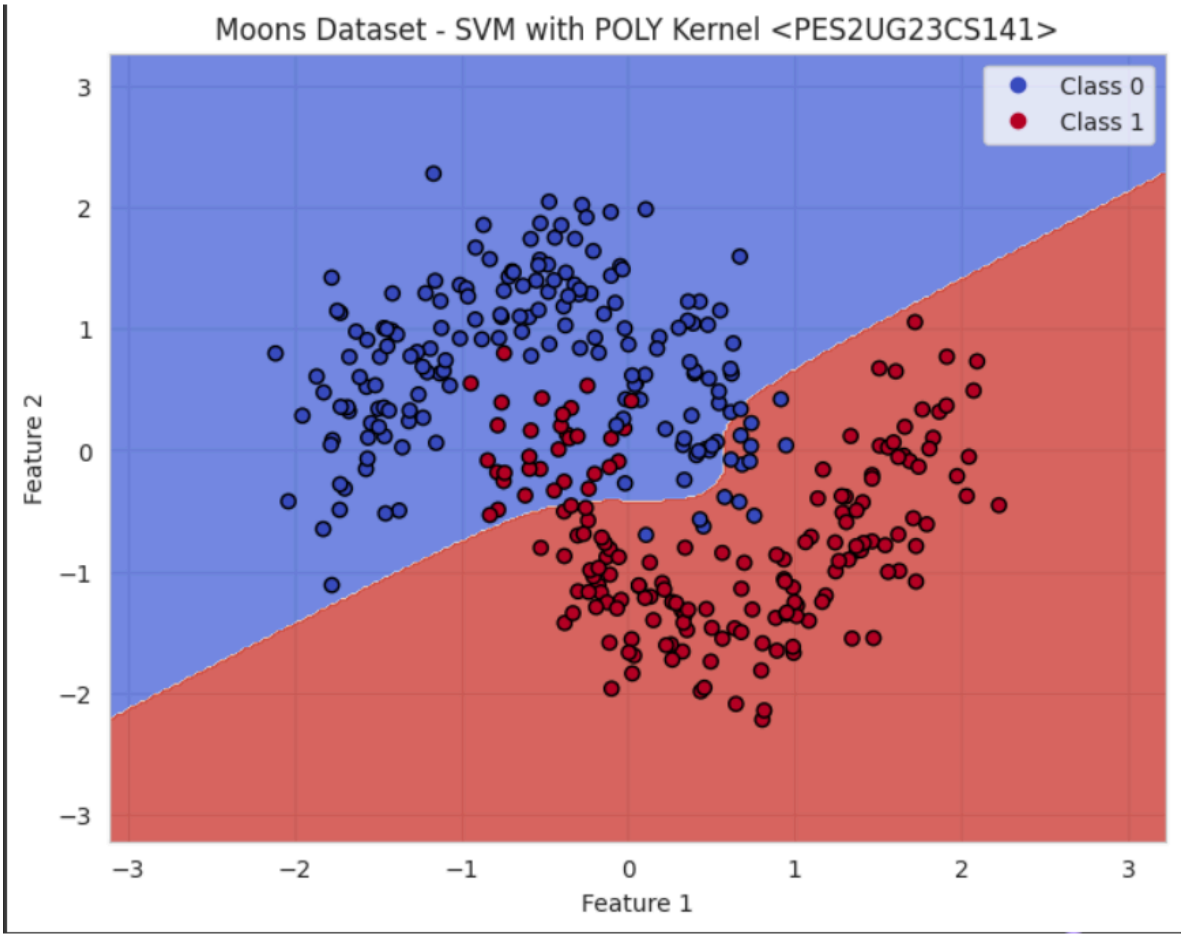
Moons Dataset - SVM with LINEAR Kernel



Moons Dataset - SVM with RBF Kernel

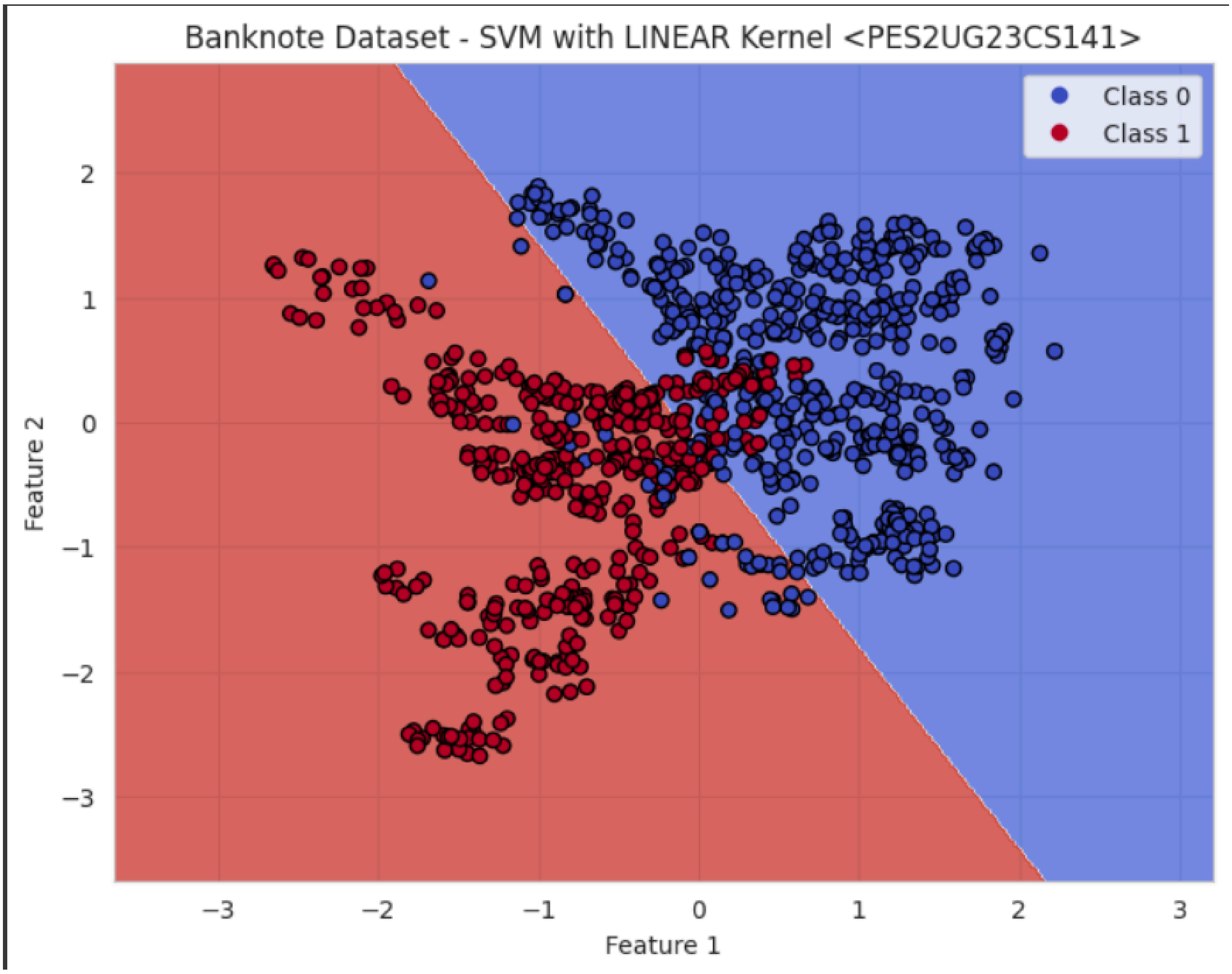


Moons Dataset - SVM with POLY Kernel

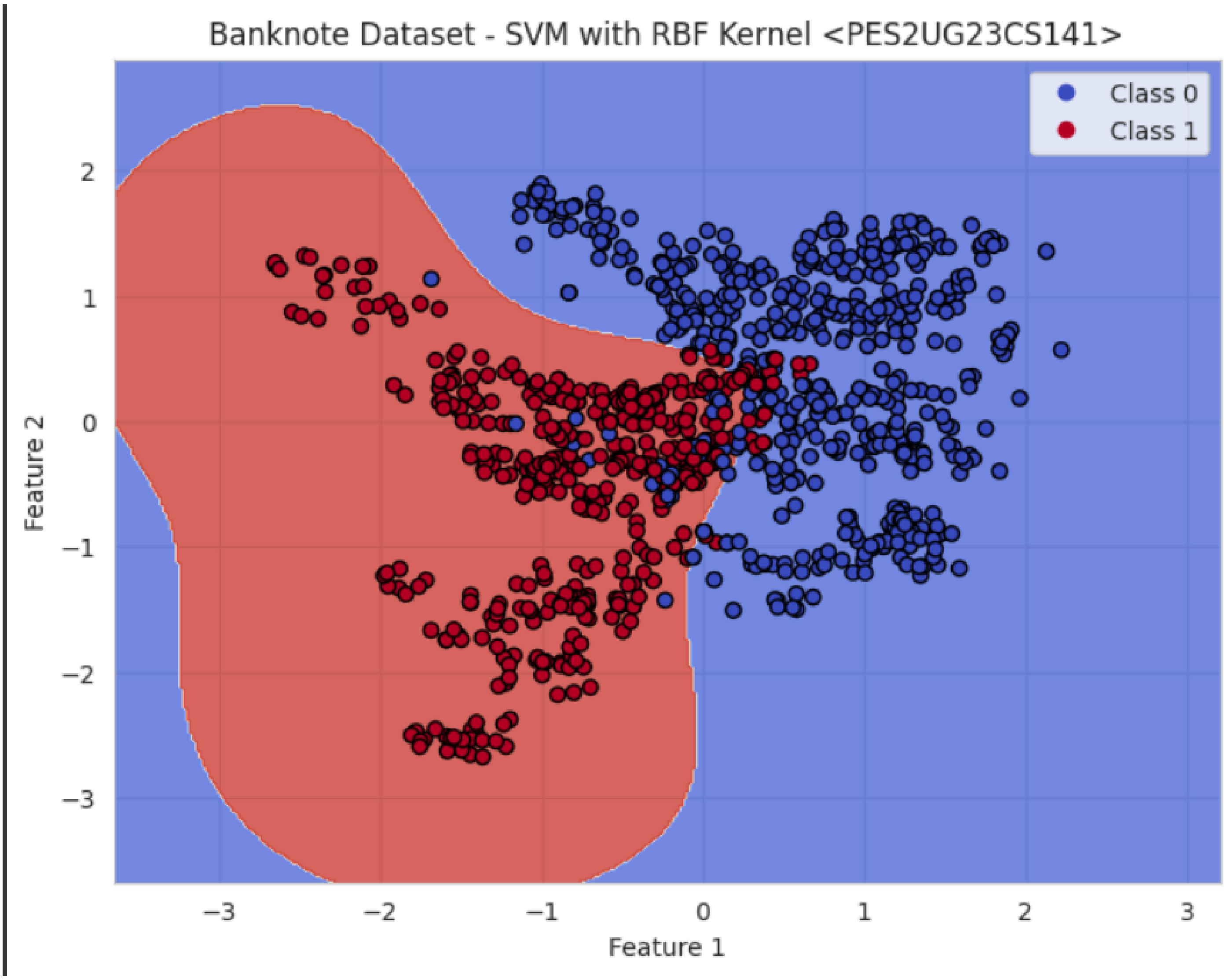


Banknote Dataset:

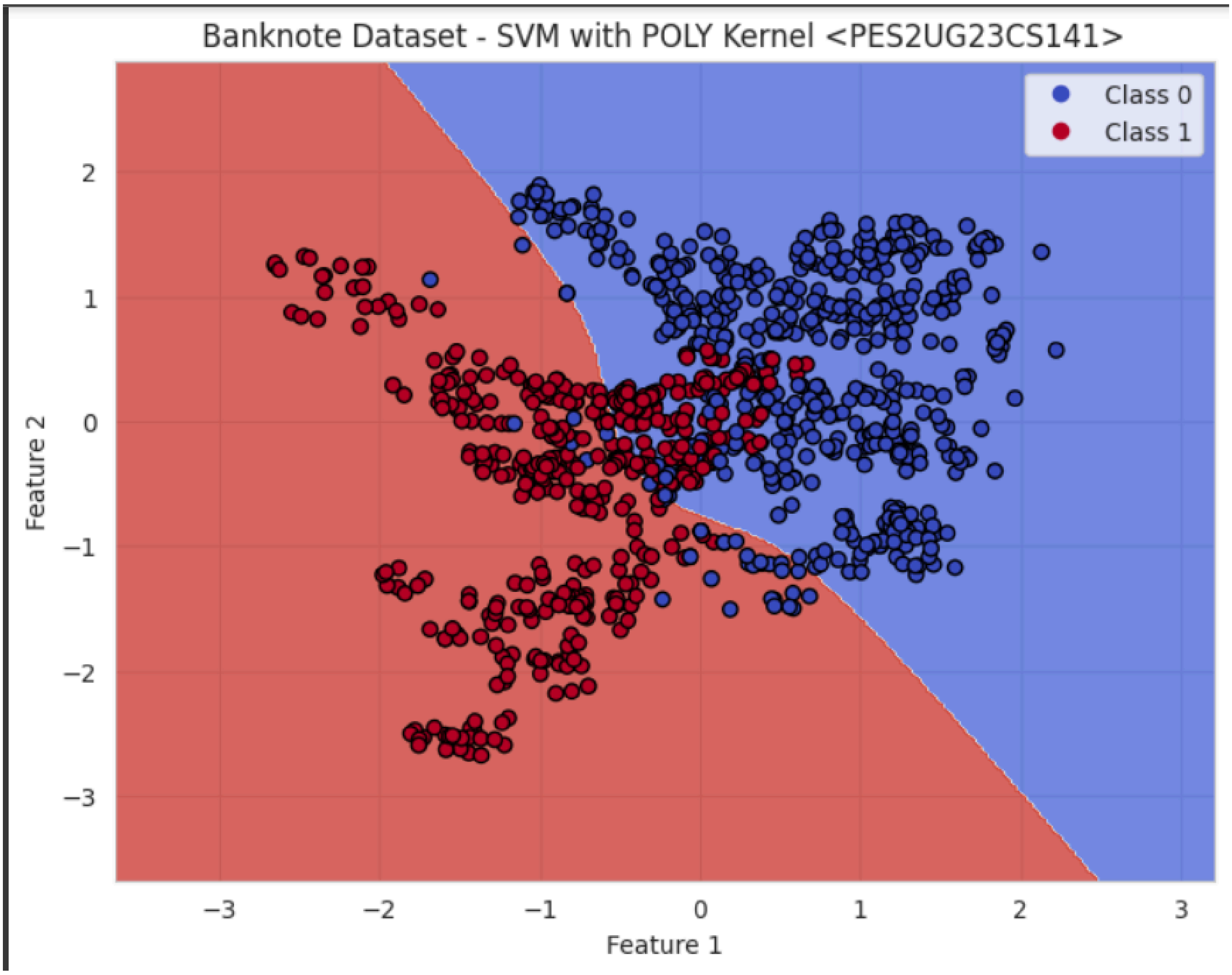
Banknote Dataset - SVM with LINEAR Kernel



Banknote Dataset - SVM with RBF Kernel

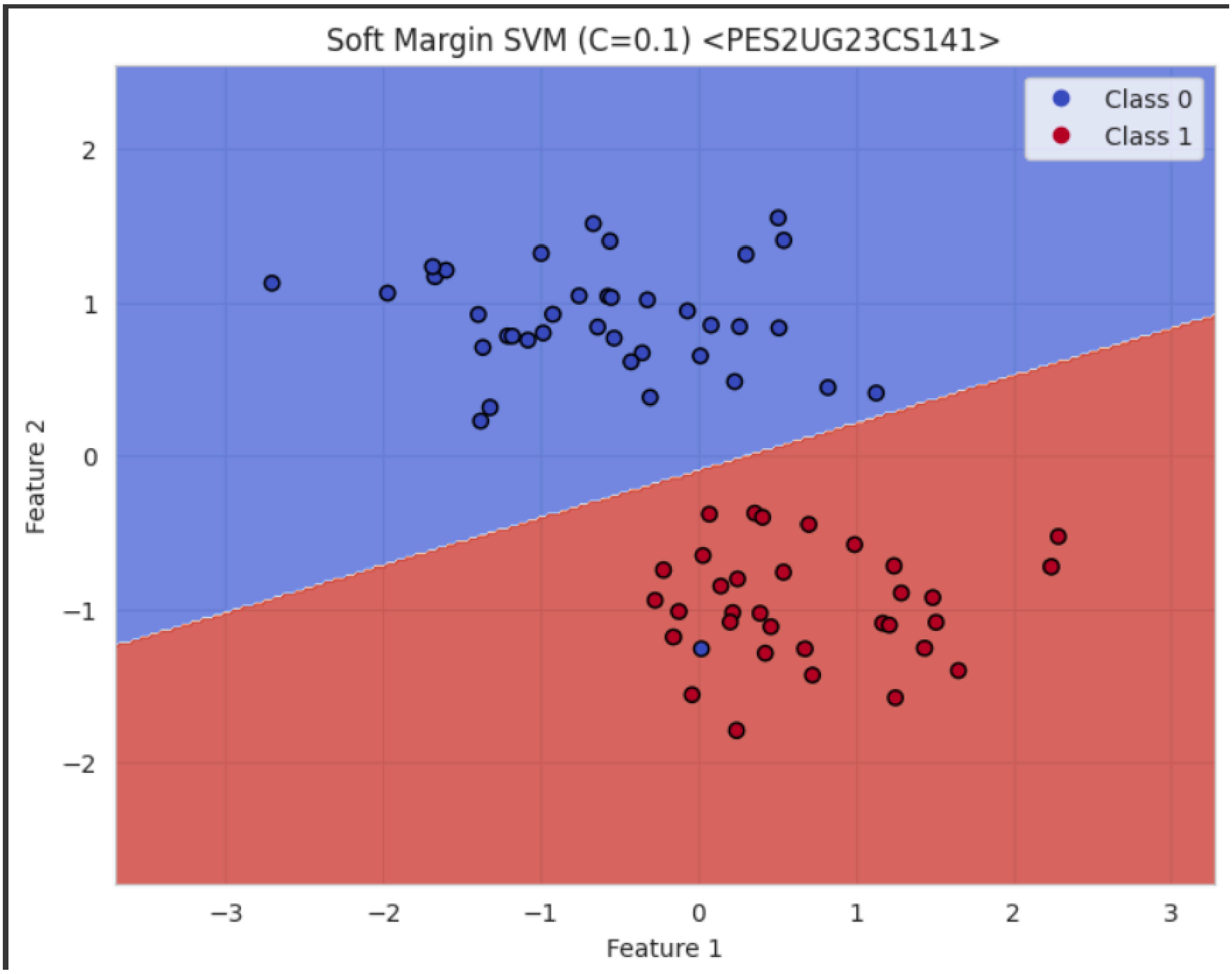


Banknote Dataset - SVM with POLY Kernel

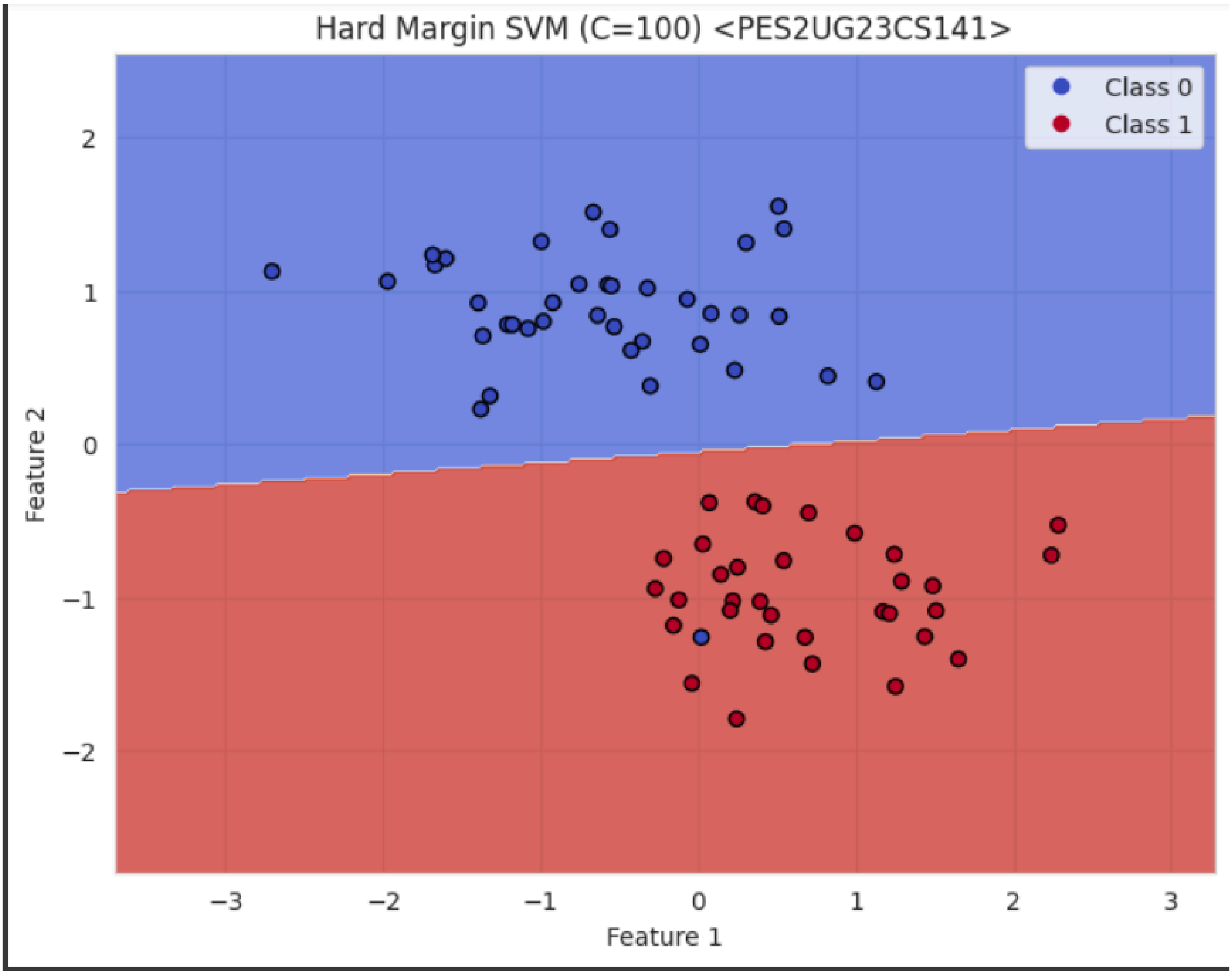


Margin Analysis:

Soft Margin SVM (C=0.1)



Hard Margin SVM (C=100)


Hard Margin SVM (C=100) <PES2UG23CS141>
Class 0
Class 1
Feature 2
Feature 1
2
1
0
−1
−2
−3
−2
−1
0
1
2
3